

WHAT IS CLAIMED IS:

1. A method for scoring a document, comprising:
identifying a document;
obtaining one or more types of history data associated with the document; and
generating a score for the document based on the one or more types of history data.
2. The method of claim 1, wherein the one or more types of history data includes information relating to an inception date; and
wherein the generating a score includes:
determining an inception date corresponding to the document, and
scoring the document based, at least in part, on the inception date corresponding to the document.
3. The method of claim 2, wherein the document includes a plurality of documents;
and
wherein the scoring the document includes:
determining an age of each of the documents based on the inception dates corresponding to the documents,
determining an average age of the documents based on the ages of the documents, and
scoring the documents based, at least in part, on a difference between the ages of the documents and the average age.

4. The method of claim 2, wherein the generating a score for the document includes scoring the document based, at least in part, on an elapsed time measured from the inception date corresponding to the document.

5. The method of claim 2, wherein the inception date corresponding to the document is based on at least one of a date when a search engine first discovers the document, a date when a search engine first discovers a link to the document, and a date when the document includes at least a predetermined number of pages.

6. The method of claim 1, wherein the one or more types of history data includes information relating to a manner in which a content of the document changes over time; and wherein the generating a score includes:
determining a frequency at which the content of the document changes over time, and
scoring the document based, at least in part, on the frequency at which the content of the document changes over time.

7. The method of claim 6, wherein the frequency at which the content of the document changes is based on at least one of an average time between the changes, a number of changes in a time period, and a comparison of a rate of change in a current time period with a rate of change in a previous time period.

8. The method of claim 6, wherein the generating a score further includes:
determining an amount by which the content of the document changes over time, and

scoring the document based, at least in part, on the frequency at which and the amount by which the content of the document changes over time.

9. The method of claim 8, wherein the amount by which the content of the document changes is based on at least one of a number of new pages associated with the document within a time period, a ratio of a number of new pages associated with the document versus a total number of pages associated with the document, and a percentage of the content of the document that has changed during a time period.

10. The method of claim 8, wherein the determining an amount by which the content of the document changes includes:

weighting different portions of the content of the document differently based on a perceived importance of the portions, and

determining the amount by which the content of the document changes as a function of the differently weighted portions of the content.

11. The method of claim 6, wherein the document includes a plurality of documents;
and

wherein the scoring the document includes:

determining a date on which the content of each of the documents last changed,

determining an average date of change based on the determined dates on which the contents of the documents last changed, and

scoring the documents based, at least in part, on a difference between the dates on which the contents of the documents last changed and the average date of change.

12. The method of claim 1, wherein the one or more types of history data includes information relating to a manner in which a content of the document changes over time; and

wherein the generating a score includes:

determining an amount by which the content of the document changes over time, and

scoring the document based, at least in part, on the amount by which the content of the document changes over time.

13. The method of claim 12, wherein the amount by which the content of the document changes is based on at least one of a number of new pages associated with the document within a time period, a ratio of a number of new pages associated with the document versus a total number of pages associated with the document, and a percentage of the content of the document that has changed during a time period.

14. The method of claim 12, wherein the determining an amount by which the content of the document changes includes:

weighting different portions of the content of the document differently based on a perceived importance of the portions, and

determining the amount by which the content of the document changes as a function of the differently weighted portions of the content.

15. The method of claim 1, wherein the one or more types of history data includes information relating to how often the document is selected when the document is included in a set of search results; and

wherein the generating a score includes:

determining an extent to which the document is selected over time when the document is included in a set of search results, and

scoring the document based, at least in part, on the extent to which the document is selected over time when the document is included in the set of search results.

16. The method of claim 15, wherein the scoring the document includes assigning a higher score to the document when the document is selected more often than other documents in the set of search results over a time period.

17. The method of claim 1, wherein the one or more types of history data includes information relating to search terms that increasingly appear in search queries over time; and

wherein the generating a score includes:

determining whether the document is associated with the search terms, and

scoring the document based, at least in part, on whether the document is associated with the search terms.

18. The method of claim 1, wherein the one or more types of history data includes information relating to queries that remain approximately constant over time but lead to results that change over time; and

wherein the generating a score includes:
determining whether the document is associated with queries that lead to results that change over time, and
scoring the document based, at least in part, on whether the document is associated with queries that lead to results that change over time.

19. The method of claim 1, wherein the one or more types of history data includes information relating to staleness of documents; and

wherein the generating a score includes:
determining whether the document is stale, and
scoring the document based, at least in part, on whether the document is stale.

20. The method of claim 19, wherein the scoring the document includes:
determining whether stale documents are considered favorable for a search query when the document is determined to be stale, and
scoring the document based, at least in part, on whether stale documents are considered favorable for the search query when the document is determined to be stale.

21. The method of claim 20, wherein the determining whether stale documents are considered favorable for the search query is based, at least in part, on how often stale documents were selected over recent documents over time for the search query.

22. The method of claim 1, wherein the one or more types of history data includes information relating to behavior of links over time; and

wherein the generating a score includes:

determining behavior of links associated with the document, and

scoring the document based, at least in part, on the behavior of links associated with the document.

23. The method of claim 22, wherein the behavior of links relate to at least one of appearance and disappearance of one or more links pointing to the document.

24. The method of claim 23, wherein the appearance of one or more links relates to at least one of a date that a new link to the document appears, a rate at which the one or more links appear over time, and a number of the one or more links that appear during a time period, and the disappearance of one or more links relates to at least one of a date that an existing link to the document disappears, a rate at which the one or more links disappear over time, and a number of the one or more links that disappear during a time period.

25. The method of claim 22, wherein the determining behavior of links associated with the document includes monitoring at least one of time-varying behavior of links associated with the document, how many links associated with the document appear or disappear during a time period, and whether there is a trend toward appearance of new links associated with the document versus disappearance of existing links associated with the document.

26. The method of claim 1, wherein the one or more types of history data includes information relating to freshness of links; and

wherein the generating a score includes:
determining freshness of links associated with the document,
assigning weights to the links based on the determined freshness, and
scoring the document based, at least in part, on the weights assigned to the links associated with the document.

27. The method of claim 26, wherein the freshness of a link associated with the document is based on at least one of a date of appearance of the link, a date of a change to the link, a date of appearance of anchor text associated with the link, a date of a change to anchor text associated with the link, a date of appearance of a linking document containing the link, and a date of a change to a linking document containing the link.

28. The method of claim 26, wherein the weight assigned to a link is based on at least one of how much a document containing the link is trusted, how authoritative a document containing the link is, and a freshness of a document containing the link.

29. The method of claim 26, wherein the scoring the document includes:
determining an age of each link pointing to the document,
determining an age distribution associated with the links based on the ages of the links,
and

scoring the document based, at least in part, on the age distribution associated with the links.

30. The method of claim 1, wherein the one or more types of history data includes information relating to a manner in which anchor text changes over time; and

wherein the generating a score includes:

identifying a change in anchor text associated with a link to the document, and

scoring the document based, at least in part, on the change in anchor text associated with a link to the document.

31. The method of claim 1, wherein the one or more types of history data includes information relating to differences in documents and anchor text associated with links to the documents; and

wherein the generating a score includes:

determining whether a content of the document changes such that the content differs from anchor text associated with one or more links to the document, and

scoring the document based, at least in part, on whether the content of the document changes such that the content differs from the anchor text associated with one or more links to the document.

32. The method of claim 1, wherein the one or more types of history data includes information relating to freshness of anchor text; and

wherein the generating a score includes:

determining freshness of anchor text associated with one or more links to the document,
and

scoring the document based, at least in part, on the freshness of anchor text associated
with one or more links to the document.

33. The method of claim 32, wherein the freshness of anchor text associated with a link to the document is based on at least one of a date of appearance of the anchor text, a date of a change to the anchor text, a date of appearance of a link associated with the anchor text, a date of a change to a link associated with the anchor text, a date of appearance of the document, and a date of a change to the document.

34. The method of claim 1, wherein the one or more types of history data includes information relating to traffic associated with documents; and

wherein the generating a score includes:

determining characteristics of traffic associated with the document, and

scoring the document based, at least in part, on the characteristics of traffic associated with the document.

35. The method of claim 34, wherein the determining characteristics of traffic associated with the document includes analyzing a traffic pattern associated with the document to identify changes in the traffic pattern over time.

36. The method of claim 1, wherein the one or more types of history data includes information relating to user behavior associated with documents; and

wherein the generating a score includes:
determining user behavior associated with the document, and
scoring the document based, at least in part, on the user behavior associated with the document.

37. The method of claim 36, wherein the user behavior relates to at least one of a number of times that the document is selected within a set of search results and an amount of time that one or more users spend accessing the document.

38. The method of claim 1, wherein the one or more types of history data includes domain-related information corresponding to domains associated with documents; and

wherein the generating a score includes:
analyzing domain-related information corresponding to a domain associated with the document over time, and
scoring the document based, at least in part, on a result of the analyzing.

39. The method of claim 38, wherein the scoring the document includes:
determining whether the domain associated with the document is legitimate, and
scoring the document based, at least in part, on whether the domain associated with the document is legitimate.

40. The method of claim 38, wherein the domain-related information is related to at least one of an expiration date of the domain, a domain name server record associated with the domain, and a name server associated with the domain.

41. The method of claim 1, wherein the one or more types of history data includes information relating to a prior ranking history of documents; and
wherein the generating a score includes:
determining a prior ranking history of the document, and
scoring the document based, at least in part, on the prior ranking history of the document.

42. The method of claim 41, wherein the scoring the document includes:
determining a quantity or rate that the document moves in rankings over a time period,
and
scoring the document based, at least in part, on the quantity or rate that the document moves in the rankings.

43. The method of claim 41, wherein the prior ranking history is based on at least one of a number of queries for which the document is selected as a search result over time, a rate at which the document is selected as a search result over time, seasonality, burstiness, and changes in scores over time for a URL-query pair.

44. The method of claim 41, wherein the determining a prior ranking history of the document includes monitoring a rank of the document over time for spikes in the rank.

45. The method of claim 1, wherein the one or more types of history data includes information relating to user maintained or generated data; and

wherein the generating a score includes:

determining whether user maintained or generated data indicates that the document is of interest to a user, and

scoring the document based, at least in part, on whether the user maintained or generated data indicates that the document is of interest to a user.

46. The method of claim 45, wherein the user maintained or generated data relates to at least one of favorites lists, bookmarks, temp files, and cache files associated with one or a plurality of users.

47. The method of claim 45, wherein the scoring the document includes:

analyzing the user maintained or generated data over time to identify at least one of trends to add or remove the document, a rate at which the document is added to or removed from the user maintained or generated data, and whether the document is added to, deleted from, or accessed through the user maintained or generated data, and

scoring the document based, at least in part, on a result of the analyzing.

48. The method of claim 1, wherein the one or more types of history data includes information relating to growth profiles of anchor text; and

wherein the generating a score includes:

determining a growth profile of anchor text associated with one or more links to the document, and

scoring the document based, at least in part, on the growth profile of anchor text associated with one or more links to the document.

49. The method of claim 1, wherein the one or more types of history data includes information relating to linkage of independent peers; and

wherein the generating a score includes:

determining a growth in a number of independent peers that include the document, and
scoring the document based, at least in part, on the number of independent peers.

50. The method of claim 1, wherein the one or more types of history data includes information relating to document topics; and

wherein the generating a score includes:

performing topic extraction relating to the document,
monitoring a topic of the document for changes over time, and
scoring the document based, at least in part, on changes to the topic of the document.

51. The method of claim 1, further comprising:
obtaining a search query, where the identified document is identified as relevant to the search query; and

generating a relevancy score for the document based on how relevant the document is to the search query; and

wherein the generating a score for the document is based, at least in part, on the one or more types of history data and the relevancy score.

52. A system for scoring a document, comprising:
means for identifying a document;
means for obtaining a plurality of types of history data associated with the document; and
means for generating a score for the document based, at least in part, on the plurality of types of history data.

53. A system for scoring a document, comprising:
a history component configured to obtain one or more types of history data associated with a document; and
a ranking component configured to:
generate a score for the document based, at least in part, on the one or more types of history data.

54. A method for ranking a linked document, comprising:
determining an age of linkage data associated with the linked document; and
ranking the linked document based on a decaying function of the age of the linkage data.

55. The method of claim 54, wherein the linkage data includes at least one link.

56. The method of claim 54, wherein the linkage data includes anchor text.

57. The method of claim 54, wherein the linkage data includes a rank based, at least in part, on links and anchor text provided by one or more linking documents and related to the linked document.

58. The method of claim 57, further comprising:
determining longevity of the linkage data;
deriving an indication of content update for a linking document providing the linkage data; and
adjusting the ranking of the linked document based on the longevity of the linkage data and the indication of content update for the linking document.

59. The method of claim 58, wherein the adjusting the ranking includes penalizing the ranking if the longevity indicates a short life for the linkage data and boosting the ranking if the longevity indicates a long life for the linkage data.

60. The method of claim 59, wherein the adjusting the ranking further includes penalizing the ranking if at least a portion of content from the linking document is considered stale over a period of time and boosting the ranking if the portion of content from the linking document is considered updated over the period of time.

61. The method of claim 54, further comprising:

determining an indication of link churn for a linking document providing the linkage data; and

based on the link churn, adjusting the ranking of the linked document.

62. The method of claim 61, wherein the indication of link churn is computed as a function of an extent to which one or more links provided by the linking document change over time.

63. The method of claim 62, wherein adjusting the ranking includes penalizing the ranking if the link churn is above a threshold.